

# **Understanding How Superintendents Use Data in a New Environment of Accountability**

Regional Educational Laboratory  
Contract #ED-01-CO-0006  
Deliverable #2003-14

*Prepared for*  
Institute of Education Sciences  
U.S. Department of Education  
Washington, DC 20209

*Prepared by*  
Kerry Englert, Ph.D.  
Senior Evaluator  
Dawn Fries, B.A.  
Evaluation Assistant  
Bryan Goodwin, M.A.  
Director of Communications  
Mya Martin-Glenn, B.A.  
Research Associate

November 30, 2003



©2003 McREL

This document has been funded at least in part with federal funds from the U.S. Department of Education under contract number ED-01-CO-0006. The content of this publication does not necessarily reflect the views or policies of the Department of Education nor does mention of trade names, commercial products, or organizations imply endorsements by the U.S. Government.

## TABLE OF CONTENTS

Introduction.....	1
Six Characteristics of Effective Accountability Systems.....	2
Method.....	4
Survey Development.....	5
Sampling.....	7
Analyses.....	7
Findings.....	7
Demographics.....	7
Accountability and Data Use.....	8
Comparisons of Rural and Non-Rural Districts.....	11
Comparisons of High- and Low-SES Districts.....	13
Conclusions and Next Steps.....	15
References.....	16

## INTRODUCTION

The purpose of the study is to better understand the assessment and accountability practices and policies superintendents are implementing and to examine whether those policies simply comply with the No Child Left Behind Act or if they go also are effective practices that further school and student improvement. This study generated descriptive information about the needs schools and districts have for data, how schools and districts are using data to guide classroom practice, and if there are differences in data use based on district location or socio-economic status. This report, which focuses on superintendents, is the first in a series that will examine how various stakeholder groups use assessment data and accountability systems.

In order to understand superintendents' use of assessment and accountability systems, the following research questions guided this study:

- What policies and practices are superintendents using to meet new accountability demands?
- Are superintendents using policies and practices that research and literature have identified as being effective?
- Do policies and practices differ for rural and non-rural schools and for schools of high and low socio-economic status?

The current attention to accountability in education has resulted in an increased focus on assessing and monitoring student achievement. Many education researchers and scholars cite both positive and negative impacts of these changes. On the positive side, increased accountability is credited with focusing schools on learning outcomes, emphasizing efforts that help all students learn, providing staff with information to adjust curricula or staffing, and informing parents about how well their children and their schools are performing. Others cite negative impacts such as a narrowing of the curriculum based solely on test content, increased retention rates, increased numbers of students placed in special education programs, and an increased focus on test performance rather than on factors such as school safety, dropout rates, and discipline.

Goodwin, Englert, and Cicchinelli (2002) reviewed the literature on accountability systems and identified 12 frequently cited components and characteristics that are essential to examine when evaluating accountability systems<sup>1</sup>. Six of these were determined to be most relevant to superintendents and were selected and used as the foundation for constructing a survey of superintendents. The selected components and characteristics are discussed in more detail in the following section.

---

<sup>1</sup> The 12 essential components and characteristics identified by Goodwin et al. were (1) clear standards and expectations; (2) high-quality assessments aligned with standards; (3) multiple measures; (4) high expectations for all students; (5) readily understandable to the public; (6) diagnostic uses for data; (7) sanctions and rewards linked to results; (8) flexibility and fairness to allow for local differences and creativity; (9) alignment of resources, support and assistance for improvement; (10) balanced, comprehensive design; (11) stakeholder support/engagement; and (12) fairness provisions.

## SIX CHARACTERISTICS OF EFFECTIVE ACCOUNTABILITY SYSTEMS

The six characteristics selected from the framework for evaluation of accountability systems (Goodwin et al., 2002), as relevant to superintendents, are (1) high-quality assessments aligned with standards; (2) alignment of resources, support, and assistance for improvement; (3) sanctions and rewards linked to results; (4) multiple measures; (5) diagnostic uses for data; and (6) readily understandable to the public. Each of these is discussed in this section, along with relevant findings from the literature.

***High-quality assessments aligned with standards.*** Although many districts use multiple data sources in order to more accurately account for and gauge their progress, student achievement, as measured through statewide assessments, is becoming synonymous with accountability. Assessments have become the primary tool for gauging students' success as well as the success of teachers, schools, and districts. Because of the central role that assessments play in accountability systems, it is necessary to ensure that they are of high quality. After specifying clear academic standards, authors of most accountability systems contend that student progress on standards needs to be measured using assessments aligned to those standards (Guth et al., 1999; Baker, Linn, Herman, & Koretz, 2002; Buckendahl, Impara, & Plake, 2002). Moreover, most accountability standards and models also place a heavy emphasis on the quality of these examinations. Thus, the most relevant aspects of quality for the surveyed population were deemed to be alignment of the assessment to state standards, technical quality, and the provision of useful data to stakeholders.

***Alignment of resources, support, and assistance for improvement.*** Many models of accountability include the need to align resources and support with the goals of the system. Grissmer and Flanagan (1998) note that "recent research is now supporting the hypothesis that resource levels can make significant differences in achievement, and that disadvantaged students probably benefit more from increased educational resources" (p. 24). As a result, "the distribution of resources among schools and school districts needs to be perceived as fair and equitable for an assessment and accountability system to function effectively" (p. 24).

In addition to equitable funding, researchers have argued that states must help schools develop the capacity to make necessary improvements and the flexibility to direct the funds toward specific programs that target areas needing improvement. Grissmer, Flanagan, Kawata, and Williamson (2000) found that lower SES states can show significant gains in student achievement "through modest increases in resources, if allocated to specific programs" (p. 101). WestEd researchers made the same assertion, noting that schools identified as not meeting performance goals should have some avenue of assistance to help them improve (Guth et al., 1999). Fuhrman (1999) pointed out that although research "shows that new accountability systems can be motivating" (p. 8), "the mere imposition of a new accountability system ... does not unleash some hidden capacity" (p. 9). In short, though accountability systems may create more *will*, they often fail to provide schools with the *way* (i.e., tools and strategies) to succeed.

In light of such concerns, Sirotnik and Kimball (1999) argue that "the accountability system must include support for and monitoring of substantial, long-term professional development opportunities for teachers and administrators to inquire into their disciplines and to review and revise their pedagogical content knowledge and teaching and leadership skills (including evaluation and assessment)" (p. 211). Proper training and skills in assessment practices are vital because teachers can spend from a third to a half of their time engaged in assessment-related activities (Stiggins, 1988, as cited in Trevisan, 1999). Therefore, administrators have an important responsibility to teachers as well as other administrators in providing leadership on assessment issues. A study by Trevisan (1999) found that only "14 states have assessment requirements for superintendents" (p. 9). Even with these requirements in place, there are no guidelines for specifying or developing the needed competencies.

***Sanctions and rewards linked to results.*** There are many examples of accountability models that use sanctions and rewards to spur school improvement efforts, despite the lack of clear evidence that either strategy is effective. In a report that examined the gains in elementary and junior high school test scores in Texas and North Carolina, Grissmer and Flanagan (1998) found that both states “have financial rewards for schools based on performance, and have the power to disenfranchise school districts and remove principals based on sustained levels of poor performance” (p. iv). Walberg (2002) argued for the use of sanctions and rewards such as those evident in Texas and North Carolina — sanctions and rewards that move beyond publicly highlighting school successes and failures — because “simply publishing results appears insufficient for progress. ... Schools of choice risk closing if they attract no students. Analogous thinking dominates much of society. Why not schools?” (p. 159). Accordingly, Education Week’s (2002) ratings of state accountability systems includes whether the state “holds schools accountable for performance” (p.12) through rewards and sanctions, including closure, reconstitution, transfers, and withholding of funds.

Sanctions and rewards are listed in the National Center for Research on Evaluation, Standards, and Student Testing’s (CRESST) *Standards for Educational Accountability Systems* with the advisement that they should start out broad and diffuse, then “move to specific consequences for individuals and institutions as the system aligns” (Baker et al., 2002, p. 5). This allows stakeholders the opportunity to make the changes necessary to meet expectations. Baker et al. caution, however, that “if test data are used as a basis of rewards or sanctions, evidence of technical quality of the measures and error rates associated with misclassification of individuals or institutions should be published” (p. 3). They also stress the importance of having an appeal procedure in place so that individuals have a way to address possible extenuating circumstances.

Sirotnik and Kimball (1999) disagree with linking progress to sanctions and rewards in any fashion. In their view “the accountability system must not be punitive, either to students or to their teachers and schools” (p. 213). After reviewing principles of learning theory, they found that punishment did not seem to be effective in changing behavior. It would be more productive, Sirotnik and Kimball assert, to support districts and schools that are not making adequate progress than to punish them.

***Multiple measures.*** Many accountability guidelines call for the use of multiple measures in order to ensure a more complete and accurate assessment of students, teachers, and schools (Sirotnik & Kimball, 1999; Baker et al., 2002). Sirotnik and Kimball argue that an “accountability system must not be driven by a single indicator (e.g., test scores) and simplistic formulas for rewards or sanctions based on that indicator” (p. 211). Similarly, researchers at CRESST maintain that “decisions about individual students should not be made on the basis of a single test” (Baker et al., 2000, p. 3). They suggest that it “is important to consider other student outcome data such as attendance, mobility, and rates of retention in grade, dropout and graduation” (p. 2). Other measures that can be useful in evaluating schools include school safety data, parent feedback, expulsion rates, principal feedback, and student performance (as measured by course grades and homework).

***District and school personnel data usage.***<sup>2</sup> Another key purpose of accountability systems is to provide schools with information they can use to guide their decision making. Although focusing on test results can help to focus educators on what’s important — student learning — in some districts there has been a tendency to focus only on test scores and ignore other data that schools could use to guide their improvement efforts. By reporting only test scores, schools will be in a position of knowing they need to

---

<sup>2</sup> In the Goodwin, Englert, and Cicchinelli (2002) report, this component is entitled “diagnostic uses for data.”

improve but not knowing how. Fuhrman (1999) found that many schools labeled as inadequate did not use test results diagnostically. The label created a state of urgency in the school, but educators had little guidance about where to channel their urgency.

In order to provide schools and teachers with diagnostic information, the data must be specific enough to be useful in identifying and correcting problems. CRESST's standards, for example, call for accountability systems to "include data elements that allow for interpretations of student, institution, and administrative performance" (Baker et al., 2002, p. 2). Including data from all levels facilitates the process of identifying areas of improvement in each area (American Education Research Association, American Psychological Association, & National Council on Measurement in Education, 1999). The key, according to Reeves (2002), is providing school communities with sufficient data so that they can determine what works. Accordingly, it's important to step back for a moment and consider all the data that states, districts, and school might collect and the purposes these data might serve.

*Informative to parents and the community*<sup>3</sup>. Another commonly identified purpose of accountability systems is to inform students, parents, and teachers about student progress and status. In order to do that, many argue, accountability systems should produce data and publish results that are meaningful to all interested parties. In its evaluation of state accountability systems, the Princeton Review (2002), for example, checks to see if the "performance data [are] shared with the public along with explanation and contextual detail appropriate for a general audience" (p. 6).

According to Walberg (2002), one principle that encourages the effective use of accountability system data is "user-friendliness" (p. 158). Reports about student and school progress should be useful and understandable to all interested parties regardless of their level of technical knowledge. It would be most effective to report the data in various forms so that parents, school board members, and educators can readily make sense of the information. "What isn't as useful," argues Walberg, "is a mass of undigested numbers often reported by states and districts in large, unwieldy books of computer printouts" (p. 158).

Although it may be helpful to members of the public to combine data into a single or small set of numbers, CRESST researchers argue that when this approach is taken, states need to make explicit "the weighting of elements in the system, including different types of test content, and different information sources" (Baker et al., 2002, p. 2). They also recommend that "results should be made broadly available to the press, with sufficient time for reasonable analysis and with clear explanations of legitimate and potential illegitimate interpretations of results" (p. 5).

## METHOD

In order to develop answers for the three research questions identified at the beginning of this report, a survey was developed and sent to a sample of superintendents. The section that immediately follows describes the development of the survey including the incorporation of the six characteristics described above. Following this, we explain how we selected the sample of superintendents.

---

<sup>3</sup>In the Goodwin, Englert, and Cicchinelli (2002) report, this component was entitled "readily understandable to the public."

## SURVEY DEVELOPMENT

In addition to the characteristics, our review of literature identified several important components of assessment and accountability systems, which were included in the development of the survey. Thus, the survey was based on two aspects: background information on the district and superintendent representing the additional components and the six characteristics. Prior to distributing the survey to superintendents, the survey was pilot-tested by a small group of former superintendents. The pilot test resulted in suggestions for clarifying the wording of existing items and including additional items. A final survey was developed and reviewed by the project team based on the reviewers' comments.

The survey measured superintendent perceptions on accountability through closed-ended questions and examined detailed contextual information regarding how superintendents were using data and encouraging their staff to use data via open-ended items. Using both of these formats resulted in a more complete picture of the policies and practices superintendents were advocating. Scales were developed from the closed-ended items to better interpret the findings.

**General information.** The survey gathered general information about the respondent and the district in which they work.

*Demographics.* The items included in this category were relevant for answering the third research question as to whether policies differ by type of district. These items included the number of schools and the number of students in the district, district location (rural, urban, or suburban), per-pupil expenditure, percentage of minority students, and percentage of students receiving free and reduced-price lunch.

*Superintendent Background.* Superintendents were asked to identify their highest degree, years of experience in K–12 education in general and as a superintendent, and their length of time as a superintendent in their current district. These items were used to determine if there were differences in practices between superintendents with different backgrounds.

*District Proficiency.* Superintendents were asked their perceptions of how well schools were performing on the state assessment in general, and how schools were performing when other data were considered in conjunction with state assessment results. Superintendents also were asked about the degree to which there were differences in achievement among subpopulations within their district and within schools in their district. These items were used to create a district proficiency scale.

**Accountability-specific scales.** For the effective characteristics described earlier, selected items were combined to create cohesive scales about each one in order to provide reliable and valid results. These scales were (1) alignment of resources, support, and assistance for improvement; (2) high-quality assessments aligned with standards; (3) informing parents and community; (4) district and school personnel data use; and (5) applying sanctions and rewards. The applying sanctions and rewards scale incorporates the multiple measures characteristic. The open-ended responses could not be scaled but provided detailed and contextual information.

*Alignment of resources, support, and assistance for improvement.* This scale was intended to measure whether adequate resources exist to support changes required by NCLB at the district and classroom levels. This scale included items to identify the adequacy of resources available for districts. Specifically, superintendents were asked about the adequacy of resources to support teachers in using data and to support

principals in improving their schools. Flexibility in the ability to align resources also was included.

*High-quality assessments aligned with standards.* Because the state assessment system has become the focal point of accountability under NCLB, items were included that measured superintendents’ perceptions concerning the quality of their state assessment. It could be reasoned that if superintendents considered their state assessment system to be of high quality in terms of alignment to standards or providing usable results, they would be more inclined to consider the results in their decision making. Thus, items in this category included how well the state assessment aligned to standards, was informative to parents, provided teachers with access to results, was comprehensive, and provided diagnostic data.

*Informing parents and community.* This scale measured the degree to which superintendents used data to inform community members, in particular parents, about the progress of schools in the district. This included asking about specific systems or policies in place to ensure that results were communicated to parents and other community members.

*District and school personnel data use.* An important consideration is that, in addition to using data themselves, superintendents support their staff in using data so that changes can be made at the school and classroom levels. Items in this scale included the extent to which superintendents encouraged staff to incorporate data into their practices by providing professional development, developing policies and expectations for data usage, and asking principals to monitor their schools with data-driven goals.

*Applying sanctions and rewards.* This scale examined one of the main purposes of the survey, which was describing superintendents’ practices regarding assessment and accountability data. These items examined how and to what extent superintendents incorporated data-driven methods into their practices of monitoring and accrediting schools, and evaluating their personnel. The scale also included the degree to which superintendents used multiple measures in their monitoring of schools. Open-ended items were used to understand the breadth and depth of superintendents’ practices.

Reliability analyses were conducted on the closed-ended items to determine how well individual items within a scale correlated with each other, resulting in the reliability coefficients shown in Table 1. The reliability was adequate to warrant using the scales in further analyses and reporting.

**Table 1. Reliability of Survey Scales**

<b>Variable</b>	<b>Reliability</b>
District proficiency	.65
Alignment of resources, support, and assistance for improvement	.82
High-quality assessments aligned with standards	.87
Informing parents and community	.58
District and school personnel data usage	.74
Applying sanctions and rewards	.76

## **SAMPLING**

The seven state Departments of Education in the Central Region were asked to participate in this study. Four of the seven states agreed to participate and were asked to identify 20 districts that would result in a representative sample in terms of low, middle, and high performing; rural, suburban, and urban; varying amounts of per-pupil expenditures; and varying percentages of minority/ELL students<sup>4</sup>.

Of the 80 superintendents initially contacted, approximately 25 percent returned a completed survey. Follow-up methods resulted in a total of 48 completed surveys and an overall response rate of 60 percent. Examination of the 22 non-respondents revealed no systematic pattern in terms of location or size of district.

## **ANALYSES**

Several different analyses were conducted on the survey data. First, because the primary goal of this study was to gather descriptive evidence about superintendents' use of data, means, frequencies, and standard deviations were computed on individual items and overall scales. Second, an analysis was completed on the open-ended response items. Open-ended questions were used throughout the survey to obtain a more comprehensive picture of how superintendents are using data for accountability purposes and in what contexts. The comments for each item were analyzed for common themes. A theme was considered to be any specific problem, concept, or opinion that two or more respondents included in their written answers to a particular question. The comments were then coded by theme; a single comment could be coded under multiple themes. The open-ended results are discussed in conjunction with the results of the corresponding closed-ended items. Finally, comparisons were made between rural and non-rural districts, given concerns about rural schools having adequate resources to meet the demands of NCLB, and similarly between high- and low-SES districts.

## **FINDINGS**

The findings from the analysis of survey data are organized into four categories: demographics, accountability and data use, comparisons of rural and non-rural districts, and comparisons of high- and low-SES districts.

### **DEMOGRAPHICS**

The demographics of responding districts were analyzed to determine if they appropriately represented the states from which they were selected. This helped ascertain if there was a response or sampling bias. Based on data from the Common Core Data Set from the National Center for Educational Statistics, 31 to 77 percent of schools in the participating states were identified as rural. In the actual sample, 63 percent of districts identified themselves as rural. This was within the range from the actual state values. With regard to per-pupil expenditures in the participating states, the actual 2002 range was from \$6,244 to \$6,906. The average for the sample was approximately \$6,000, which indicates either a drop in funding from 2002 to 2003 or, more likely, slightly lower levels of funding in the sampled districts relative to the states' averages.

---

<sup>4</sup> Having the commissioners of education select the sample may have resulted in sampling bias; however, because the commissioners were most familiar with their states and districts, it was determined that they would be best qualified to purposively select appropriate districts.

In terms of student demographics, 2002 data for the surveyed states showed that from 28 to 35 percent of students were eligible for free or reduced-price lunch. The reported data from superintendents indicated a much higher figure of 43 percent for free and reduced-price lunch. Common Core Data for 2001 showed that the state averages for the percentage of minority students ranged from 14 to 33 percent, with most states in the upper end of that distribution. The survey indicated that the participating districts have a higher percentage of students eligible for free or reduced-price lunch, but are similar to the state averages in terms of percentage of minority students. Tables 2 and 3 report these data.

**Table 2. Sample Frequencies**

Variable	Percent N=48
State	
Colorado	27.1%
Kansas	29.2%
Missouri	16.7%
South Dakota	27.1%
District Location	
Rural	62.5%
Suburban	25.0%
Urban	12.5%

**Table 3. District Demographics**

Variable	Mean N=48	SD
Number of schools in district	19.9	32.6
Number of students in district	10,057.0	18,218.1
Per-pupil expenditure	5,999.8	915.1
Percentage of minority students	27.4	26.3
Percentage of student receiving free or reduced-price lunch	42.8	18.7
Years in K–12 education	28.9	8.0
Years as a superintendent	9.7	6.8
Years as a superintendent in current district	5.6	4.0

## ACCOUNTABILITY AND DATA USE

The next step in the analysis was to examine the results of the closed- and open-ended items that corresponded to each scale. Table 4 includes the mean and standard deviation for the combined closed-ended items that made up each scale. A 5-point scale was used, ranging from 1 = low to 5 = high. The following sections describe the results.

***District proficiency.*** Each superintendent was asked to rate the performance levels of the schools in their district with regard to overall proficiency on the state assessment, overall proficiency considering additional sources of data (i.e., dropout rates, expulsion rates, principal feedback, and other assessment results), and consistency in performance across schools and between demographic groups (i.e., race, socio-economic status, special education, and English Language Learners). With regard to the level of

district proficiency, most superintendents thought that their districts were fairly proficient and rated them 3.32 on average. Superintendents were asked to rate the performance of their schools, once when considering multiple sources of data, and second, when considering only performance on the state assessment. Superintendents rated their schools' proficiency higher in the first case compared to the second. Many also indicated that student performance in schools in their districts was improving over time, but noted a degree of variability in achievement among their schools and among different subsets of students in those schools.

**Table 4. Descriptive Statistics for Scales**

Variable	Mean N=48	SD
District proficiency	3.32	.51
Alignment of resources, support, and assistance for improvement	3.15	.75
High-quality assessments	3.47	.63
Informing parents and community	4.07	.49
District and school personnel data usage	3.67	.60
Applying sanctions and rewards	3.98	.59

***Alignment of resources, support, and assistance for improvement.*** Responses to this scale on resources and support were the lowest of all the scales that were measured in the survey (3.15). Superintendents reported that they lacked sufficient resources overall and, specifically, that they needed additional resources to support their principals and teachers in improving their schools and classrooms.

When superintendents were asked to list the top three issues that their districts were facing, the most common response was that they lacked the resources needed to train teachers how to analyze and use data to improve their classroom practices. The following comment captures this sentiment:

[I need] adequate and productive time for teachers and administrators to (1) fully analyze and develop strategies to address student areas of needs, (2) [develop a] complete understanding on how to manipulate and or disaggregate data to identify student and curriculum needs, [and] (3) use the data to help restructure learning and instructional time to best meet all student needs.

Overall, most superintendents also cited lack of funds as a major root problem. Many also responded that they needed more support from their states to assist their improvement efforts regarding state assessment results, specifically in order to improve professional development. One superintendent stated, “[We need] good professional development on using assessments to formulate instruction and leadership on sound instructional practices and the research base that supports them, especially for subgroup populations.”

***High-quality assessments aligned with standards.*** This scale included items focused on superintendents' perceptions concerning the quality and utility of their state assessments, which is critical given the increased focus on state assessments as a primary accountability measure. The average response was 3.47. Superintendents rated the quality of assessments lower when asked about how useful the assessment data were to parents, and the extent to which the assessment provided diagnostic data to teachers.

One common theme in the open-ended questions was the need for better state assessment reports and more timely feedback. For example, one superintendent expressed a need for “more online assessments

with timely feedback for staff, better communication regarding assessment results and a more simplified report for schools and districts and additional funds earmarked for technology and support.” The importance of improving the technical quality of state assessments, such as their reliability and validity, was also a theme.

***Informing parents and community.*** Mean values for this variable were the highest at 4.07, indicating that superintendents were using and encouraging the use of data to inform parents and the community. Ninety percent of superintendents indicated that they had a policy or expectation that results be communicated to parents. Some policies were very specific. For example, one superintendent outlined a comprehensive communication system:

All testing data on the individual student is sent by mail to the parent(s) each year along with an explanation of the data. Teachers and counselors during parent conferences will cover the information and aid in its interpretation. According to the summer school policy, this begins in October with more intensive information and follow-up through the second semester.

Conversely, other respondents’ policies consisted of simply mailing results home to parents.

Regardless of written policies or expectations, several respondents cited concerns around improving parent and community understanding about assessments and needing to educate parents about student testing.

***District and school personnel data usage.*** The scores on this variable (3.67) indicate that superintendents are encouraging their staff to use data at the school and classroom levels. About three-quarters of the superintendents indicated that there was a district policy or expectation that teachers use data to guide their classroom practices. Given the recent attention placed on achievement data, one might expect all superintendents to have at least an expectation that their staff would use data. Most districts are providing professional development opportunities to help teachers use data in their classrooms (4.27), but superintendents were less likely to think that schools were developing a culture of using data to improve classroom practices (3.27).

When superintendents were asked about resources to support data use, 75 percent cited having software, access to online data, and available or dedicated staff to help schools and themselves with this task. One superintendent discussed a very thorough program:

We assess all students at the beginning and end of each year, and teachers are expected to assess throughout the year as well to monitor progress. These data, along with other forms of assessment and classroom performance, dictate the individualized portions of the students’ instruction.

Many superintendents responded that their districts were providing some professional development to help teachers use data, but the nature of the professional development varied widely from short-term, single sessions to embedded and continuous programs. For example, one respondent with a complex district professional development program replied:

We provide weekly staff inservice in various areas such as teaching strategies, assessments, etc. We also encourage and pay for workshops and seminars for teachers and paraprofessionals to attend. We have also paid for college class work if it pertains to reading and/or math to get all of the staff highly qualified.

Another district simply utilized its state department of education as its only form of data-driven professional development.

**Applying sanctions and rewards.** Superintendents indicated that they use data to monitor schools and apply rewards and sanctions (3.98). They expressed the opinion that monitoring student achievement contributes to school success. Most superintendents monitor their schools using multiple data sources. Over half of the superintendents indicated that they used between five and eight of 13 example data sources to monitor their schools. The most frequently cited sources were dropout rates, parent feedback, district-created assessments, and other standardized assessments. One superintendent outlined his or her practices:

I utilize this [assessment data] in my overall assessment of principals although it is generally only one piece of the evaluation. Consistent and significant decline or consistent lack of growth could be a reason I would remove a principal although this has never been the case. I have had to put principals on notice that immediate improvement is needed.

Superintendents also were likely to monitor the progress of schools using different measures. Approximately 70 percent of the superintendents used three or more measures to gauge their schools' progress. Of these, the most frequently cited measures were adequate yearly progress, percentage proficient by cohort, and percentage proficient by grade level. Although superintendents indicated a high degree of sophistication in their efforts to monitor school and student achievement, they did not frequently use data to apply rewards or sanctions (only 30 percent of superintendents had a district policy of linking assessment results to rewards or sanctions).

**Relationships among scales.** Although it is important to note the mean value for each scale, relationships also were examined among the scales to look for patterns in responses. While many of the scales were independent of each other, one interesting finding emerged. Among the three scales that examined data usage (superintendents using data to monitor their schools, superintendents using data to communicate with parents and the community, and district and school personnel data usage) significant correlations were found between the scales that ranged from 0.45 to 0.55<sup>5</sup>. This indicates that superintendents who used data were likely to do so in multiple ways and contexts.

## COMPARISONS OF RURAL AND NON-RURAL DISTRICTS

Because McREL serves a region that is predominantly rural, we examined disaggregated data by district location (rural versus urban/suburban). Descriptive statistics, effect sizes and tests of significance were used to compare the two groups.

**Demographics.** When the sample was analyzed to compare rural and non-rural districts, obvious differences were apparent in district size in terms of numbers of schools and students (Table 5). In addition, rural districts were found to have a much higher percentage of students who were eligible to receive free or reduced-price lunch compared to non-rural districts.

---

<sup>5</sup> Correlations were significant to 0.01.

**Table 5. Comparisons of Demographics for Rural and Non-Rural Districts**

Variable	Rural Superintendents N=30		Urban/Suburban Superintendents N=18	
	Mean	SD	Mean	SD
Number of schools in district	6.5	8.0	41.5	45.4
Number of students in district	2259.7	3846.4	23272.1	25028.3
Per-pupil expenditure	6,111.5	970.6	5,815.9	809.6
Percentage of minority students	21.3	26.1	27.9	23.6
Percentage of student receiving free or reduced-price lunch	48.3	16.5	34.6	20.0
Years in K–12 education	29.8	6.1	29.7	7.9
Years as a superintendent	9.3	6.8	10.0	7.4
Years as a superintendent in current district	5.9	3.8	4.7	3.3

**Comparisons on scales.** An examination of the differences on the quantitative scores for each scale found clear differences between rural and non-rural superintendents with regards to their use of data (Table 6). The two areas with significant differences were informing parents and the community and applying sanctions and rewards. These effect sizes<sup>6</sup>, - 0.78 and -0.92, are large and indicate that rural superintendents consistently rated their districts lower on all items. No systematic differences were found between the two groups across items.

Small effect size differences were seen for district proficiency, high-quality assessments, and district personnel data usage. Although rural respondents rated the proficiency of their districts higher than non-rural schools, some of these differences can be attributed to two items that asked about the variability in performance among subsets of students and across schools in the district. Urban and suburban districts are more likely to have diverse student populations, increasing the opportunity for variability across sub-groups.

Rural superintendents also were more likely to give lower ratings to both the quality of their state assessment and their staff’s use of data. It is difficult to attribute these ratings to specific reasons or differences between the two groups, and there were no patterns in item-level responses or responses to open-ended items from which to attribute the differences. It may be that the lower ratings in the area of district personnel data use are because of a lack of information in rural areas about how to incorporate data into classroom practices or locally focused accountability practices that do not include incorporating state assessment data.

---

<sup>6</sup> The effect size is the difference the two groups expressed in standardized units. An effect size is the mean of one group minus the mean of a second group, all divided by the average standard deviation. This statistic provides additional evidence when examining group differences. In this case, a positive effect size indicates that the group in the first column had a higher mean, whereas a negative value indicates that the group in the second column had a higher mean. Effect sizes are particularly useful for comparing group differences when a very small or very large sample size is used because the results of significance testing can be unduly influenced by extreme sample sizes.

**Table 6. Comparisons of Scale Scores for Rural and Non-Rural Districts**

Variable	Rural Superintendents N=30		Urban/Suburban Superintendents N=18		Effect Size
	Mean	SD	Mean	SD	
District proficiency	3.38	.55	3.22	.46	.31
Alignment of resources, support, and assistance for improvement	3.13	.78	3.18	.69	-.07
High-quality assessments	3.41	.68	3.61	.53	-.33
Informing parents and community	3.94	.55	4.31	.40	-.78*
District and school personnel data usage	3.57	.50	3.81	.73	-.39
Monitoring schools and applying sanctions and rewards	3.85	.56	4.31	.44	-.92*

\* Significant to  $p < 0.01$ . Effect sizes are not considered to be meaningful if less than .20; those between .20 and .50 are considered small; between .50 to .80, medium; and greater than .80, large (Cohen, 1988).

Even though rural and non-rural superintendents rated their resource needs at nearly the same levels, it appears as though rural superintendents might need more support in monitoring their schools, helping their staff use data at the school and classroom level, and communicating more with their communities about data. It is hard to discern one particular reason for the differences between the two groups of respondents. It is possible that rural superintendents are less inclined to use data because of the lack of highly trained assessment personnel who are available in large districts or perhaps because of a lack of connectedness with other districts or their state department due to geographical constraints. In Colorado for example, the Association of Colorado Educational Evaluators convenes meetings during which a great many of the urban assessment personnel and members of the state department discuss issues with the state assessment, how to use data more effectively, and how to improve test results. Rural districts are not excluded from this group, but because of travel constraints, they are underrepresented.

Because of the large and consistent differences in the results between the rural and non-rural schools, an additional analysis was conducted to further examine the differences between large and small districts. Districts were further divided between large (greater than 1,000 students) and small rural districts (less than or equal to 1,000 students) and large (greater than 16,000 students) and small non-rural districts (less than or equal to 16,000 students). These groups were then compared. Overall, compared to the other three groups of districts, smaller rural districts consistently rated themselves lower in areas of data use such as using data for sanctions and rewards, informing the community, and encouraging their staff to use data. This finding indicates that very small, rural districts might be well served by a professional development and technical services program that is targeted to their needs in these areas.

### COMPARISONS OF HIGH- AND LOW-SES DISTRICTS

Because of the discrepancy between rural and non-rural districts in the number of students receiving free or reduced-price lunch, the sample was split by percentage of free or reduced-price lunch into districts that were above the sample mean of 42 percent of students receiving free or reduced-price lunch and those below. This allowed us to explore whether the differences in the sample could be attributed to the population of students in those districts.

**Demographics.** The demographic data for the two samples were as is often observed when comparing high- and low-SES groups (Table 7). For example, the higher SES group had lower percentages of minorities and a higher per-pupil expenditure. Superintendents had similar levels of experience across the two groups.

**Table 7. Comparisons of Demographics for High- and Low-SES Districts**

Variable	Superintendents from High-SES Districts N=24		Superintendents from Low-SES Districts N=22	
	Mean	SD	Mean	SD
Number of schools in district	19.4	31.4	20.1	35.0
Number of students in district	10,676.0	19,231.1	9,822.5	17,804.8
Per-pupil expenditure	\$6,174.6	972.5	\$5,750.0	810.2
Percentage of minority students	12.1	10.6	37.1	29.6
Percentage of student receiving free or reduced price lunch	28.5	9.6	58.4	12.4
Years in K–12 education	29.0	6.2	29.7	8.0
Years as a superintendent	10.5	7.8	8.6	5.7
Years as a superintendent in current district	5.4	4.4	5.1	2.6

**Scale score comparisons.** When examining the scale scores, several patterns emerged (see Table 8). First, there was a large and significant difference in the level of proficiency ratings that favored the higher SES districts. Second, low-SES districts rated both the adequacy of their resources and their district personnel’s use of data much higher than high SES-districts. However, low-SES districts were less likely to use data to inform parents and the community. A reason for the higher levels of personnel data usage may be that individual districts with higher numbers of poorer students are experiencing greater pressure to improve individual student performance. Therefore, for low-SES districts, the focus on student achievement data to improve performance is greater than for high-SES districts.

**Table 8. Comparisons on Scale Scores for High- and Low-SES Districts**

Variable	Superintendents from High-SES Districts N=24		Superintendents from Low-SES Districts N=22		Effect Size
	Mean	SD	Mean	SD	
District proficiency	3.53	.44	3.07	.50	.98*
Alignment of resources, support, and assistance for improvement	3.01	.72	3.30	.79	-.38
High-quality assessments	3.57	.64	3.44	.58	.21
Informing parents and community	4.19	.64	4.01	.42	.33
District and school personnel data use	3.54	.64	3.82	.56	-.47
Monitoring schools and applying sanctions and rewards	3.98	.53	4.01	.67	-.05

\* Significant to  $p < 0.01$ . Effect sizes are not considered to be meaningful if less than .20; those between .20 and .50 are considered small; between .50 to .80, medium; and greater than .80, large (Cohen, 1988).

## CONCLUSIONS AND NEXT STEPS

This study helps to illuminate how superintendents use data in light of new state and national requirements for accountability<sup>7</sup>. Overall, superintendents are using data to a great extent in their own practices, and are encouraging their staff to use data in ways that align with the notion of effective practices in accountability. When examining the open-ended responses, superintendents gave very thoughtful and comprehensive answers to questions, indicating an overall level of sophistication about the technical aspects and practical uses of test data.

The disaggregated analyses for rural and non-rural districts and high- and low-SES districts do show relative strengths and weaknesses among the groups of superintendents. Rural superintendents, in particular, were far less likely to use data in their own practices or to encourage their staff to use data; this became even more prevalent when comparing the results of the very small rural districts to larger rural and non-rural districts. Although it is difficult to say definitively why these differences exist, it is possible that rural superintendents are less likely to network with other educators about the best ways to use data. They might also feel less pressure from the state or community to improve student assessment results if they are already performing well. These disparities in data usage among districts should be cause for concern given the increased focus on high-stakes accountability.

Though the results from this study provide interesting and useful information about superintendent practices, additional data are needed in order to obtain a more complete picture of the impact of accountability at the school and classroom levels. In subsequent months, the principals and teachers of superintendents who participated in this study will be surveyed in order to examine the interrelationships between district, school, and classroom practices and policies. This nested data structure will allow for a direct analysis among these three levels.

---

<sup>7</sup> It is important to remember when interpreting these conclusions that the respondent sample was slightly weighted to lower SES and lower per-pupil expenditures compared to state averages. This might impact the generalizability of the findings.

## REFERENCES

- American Education Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: Authors.
- Baker, E., L., Linn, R. L., Herman, J. L., & Koretz, D. (2002). *Standards for educational accountability* (Policy Brief 5). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Buckendahl, C. W., Impara, J. C., & Plake, B. S. (2002). District accountability without a state assessment: A proposed model. *Educational Measurement: Issues and Practices*, 21(4), 6–16.
- Cohen, J. (1988). *Statistical Power Analysis in the Social Sciences* (2<sup>nd</sup> Ed.). Hillsdale, NJ: Erlbaum.
- Education Week. (2002, January 10). Quality Counts 2002: Building blocks for success [special report]. *Education Week*, 21.
- Fuhrman, S. H. (1999). *The new accountability* (Policy Briefs RB-27). Philadelphia: Consortium for Policy Research in Education.
- Goodwin, B., Englert, K., & Cicchinelli, L. F. (2003). *Comprehensive accountability systems: A framework for evaluation* (REL Deliverable #2002-07). Aurora, CO: Mid-continent Research for Education and Learning.
- Grissmer, D., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas*. Washington, DC: National Education Goals Panel.
- Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us*. Santa Monica, CA: RAND.
- Guth, G. J. A., Holtzman, D. J., Schneider, S. A., Carlos, L., Smith, J. R., Hayward, G. C., et al. (1999). *Evaluation of California's standards-based accountability system*. Menlo Park, CA: WestEd, Management Analysis and Planning Inc.
- The Princeton Review. (2002). *Testing the testers 2002: An annual ranking of state accountability systems*. New York: Author.
- Reeves, D. B. (2002, March/April). *Accountability-based reforms should lead to better teaching and learning - period*. Retrieved October 27, 2003, from [www.edletter.org/current/reeves.shtml](http://www.edletter.org/current/reeves.shtml)
- Sirotnik, K. A., & Kimball, K. (1999). Standards for standards-based accountability systems. *Phi Delta Kappan*, 81(3), 209–214.
- Trevisan, M. (1999). Administrator certification requirements for student assessment competence. *Applied Measurement in Education*, 12(1), 1–11.
- Walberg, H. J. (2002). Principles for accountability designs. In H. J. Walberg (Ed.), *School accountability* (pp. 155–183). Stanford, CA: Hoover Institution Press, Stanford University Press.